HMM-based Activity Recognition with a Ceiling RGB-D Camera

Daniele Liciotti[1], Emanuele Frontoni[1], Primo Zingaretti[1], Nicola Bellotto[2], and Tom Duckett[2]
[1]*Dipartimento di Ingegneria dell'Informazione, Università Politecnica delle Marche, Ancona, Italy*
[2]*School of Computer Science, University of Lincoln, Lincoln, UK*
{*d.liciotti, e.frontoni, p.zingaretti*}*@univpm.it,* {*nbellotto, tduckett*}*@lincoln.ac.uk*

Keywords: ADLs, Human Activity Recognition, HMMs

Abstract: Automated recognition of Activities of Daily Living allows to identify possible health problems and apply corrective strategies in Ambient Assisted Living (AAL). Activities of Daily Living analysis can provide very useful information for elder care and long-term care services. This paper presents an automated RGB-D video analysis system that recognises human ADLs activities, related to classical daily actions. The main goal is to predict the probability of an analysed subject action. Thus, abnormal behaviour can be detected. The activity detection and recognition is performed using an affordable RGB-D camera. Human activities, despite their unstructured nature, tend to have a natural hierarchical structure; for instance, generally making a coffee involves a three-step process of turning on the coffee machine, putting sugar in cup and opening the fridge for milk. Action sequence recognition is then handled using a discriminative Hidden Markov Model (HMM). RADiaL, a dataset with RGB-D images and 3D position of each person for training as well as evaluating the HMM, has been built and made publicly available.

1 INTRODUCTION

The Activities of Daily Living (ADLs) are a series of basic activities performed by individuals on a daily basis necessary for independent living at home or in the community. ADLs include eating, taking medications, getting into and out of bed, bathing, grooming/hygiene, dressing, socializing, cooking, cleaning and walking. Automated recognition of ADLs is also of interest to the scientific community because of its potential applications in retail and security. Furthermore, monitoring human ADLs is important in order to identify possible health problems and apply corrective strategies in Ambient Assisted Living (AAL). ADLs analysis can provide very useful information for elder care and long-term care services.

This aspect can be observed in the recent appearance of smart environments, such as smart homes. Thanks to these advanced technologies, the assistance, monitoring and housekeeping of chronically ill patients or persons with special needs or elderly has been enabled in their own home environments, in order to foster their autonomy in daily life by providing the required services when and where needed.

By using such systems, costs can be reduced considerably, while alleviating some of the pressure on healthcare systems. However, many issues related to this technology are raised such as activity recognition, assistance, monitoring and person re-identification.

Successful research has so far focused on recognizing simple human activities or properly re-identifying a person in a particular scenario. Recognizing complex activities remains a challenging and active area of research.

For instance, dementia diseases of the elderly have a strong impact on ADLs. In fact, the aging diseases result in a loss of autonomy. Medical researches (Dartigues, 2005) have shown that early signs of diseases, such as Alzheimer, can be identified up to ten years before the current diagnostics. Therefore the analysis of possible lack of autonomy in the ADLs is essential to establish the diagnostics and give all the help the patient may need to deal with the disease.

Being able to automatically infer the activity that a person is performing is essential for many disabilities in older adults, which have been associated with functional status based on ADLs in individuals with stroke, Parkinson's disease, traumatic brain injury, and multiple sclerosis. The way to determine the autonomy of patient is to analyse his ability to execute the ADLs in his own environment. However, it can be complicated for a doctor to come and watch the patient doing these ADLs, as this would be a very time consuming task. An alternative would be to record the

patient doing ADLs with a camera.

Previous papers on activity classification have focused on using 2D video (Ning et al., 2009) (Gupta et al., 2009) or RFID sensors placed on objects and humans (Wu et al., 2007). The use of 2D videos leads to low accuracy even when there is no clutter (Liu et al., 2008). Moreover, RGB-D cameras are commonly used for the recognition of human actions (Liciotti et al., 2015). Instead, the use of RFID tags is generally too intrusive because it requires RFID tags on the people.

True daily activities take place in uncontrolled and cluttered households and offices and they do not happen in structured environments (e.g., with closely controlled background). For this reason their detection becomes a much more difficult task. In addition, each person has their own habits in carrying out tasks, and these variations in style and speed create additional difficulties in trying to recognise and to detect activities.

In this work, we are interested in reliably detecting daily activities that a person performs in the kitchen. In this context, this paper proposes an automated RGB-D video analysis system that recognises human ADLs activities, related to classical actions such as making a coffee. The main goal is to classify and predict the probability of an analysed subject action. We perform activity detection and recognition using an inexpensive RGB-D camera. Human activities, despite their unstructured nature, tend to have a natural hierarchical structure; for instance, generally making a coffee involves a three-step process of turning on the coffee machine, putting sugar in the cup and opening the fridge for milk. Action sequence recognition is then handled using a discriminative hidden Markov model (HMM). A dataset with RGB-D images and 3D position of each person for training as well as evaluating the HMM has been developed and made publicly available.

Several contributions are made by our work. First of all, our model is generic, so it can be applied to any sequential datasets or sensor types. Second, our model deals with the problem of scalability by taking into account the sequences recorded independently of the environment. Finally, our approach is validated using real data gathered from a real smart kitchen which helps to make our results more confident and our experiments repeatable.

The innovative aspects of this paper are in proposing an adequate HMM structure and also the use of head and hands 3D positions to estimate the probability that a certain action will be performed, which has never been done before, for the best of our knowledge, in ADLs recognition in indoor environments.

The remainder of the paper is organized as follows: Section 2 gives details on the state of art on human activity recognition; Section 3 describes in detail the problem formulation for the design of the HMM structure and the ADLs model that is the core of our work; the following section (Section 4) describes the collection of the data acquired and presents the RADiAL Dataset (Recognition of Activity DAily Living); the final sections present the experimental results (Section 5), and the conclusions (Section 6) with our future works in this direction.

2 RELATED WORK

Recognizing ADLs is a potential field where computer vision can really help, for example, elderly people to improve the quality of their lives (Pirsiavash and Ramanan, 2012). Several research works and several models are proposed to recognize activities with intrusive and non-intrusive approaches. Activity recognition using intrusive approaches requires the use of specific equipment such as cameras.

Previous works on detection of human activities have been developed from still images as well as videos (Maji et al., 2011) (Ryoo, 2011) (Hoai and De la Torre, 2014). Many papers have shown that modeling the mutual context between human poses and objects is useful for activity detection (Prest et al., 2012) (Koppula et al., 2013).

The recent availability of affordable RGB-D cameras, together with depth information, has enabled significant improvement in scene modeling, estimation of human poses and obtaining good action recognition performance (Jiang and Saxena, 2013) (Liciotti et al., 2014) (Sturari et al., 2016). This topic is very challenging and important because understanding and tracking human behaviour through videos has several useful applications. In (Nait-Charif and McKenna, 2004) Nait-Charif et al. developed a computer-vision based system to recognize abnormal ADLs in a home environment. The system tracked human activity and summarized frequently active regions to learn a model of normal activity and the system could then detect falling as an abnormal activity.

Activity recognition with non-intrusive systems is a complex task, and it is based on a deep analysis of the data gathered from the environment. The sensors in the environment record the events about the state and any changes that happen within it. Each sequence of events is associated to a particular activity. The same person can perform an activity in several ways. This variation in the behaviour of a person leads to the generation of a set of patterns that char-

acterize this person. In this light, the variability in the person's behaviour and activity, detecting interesting patterns among many others, is a task of great importance for understanding the general behaviour of the person (Ali et al., 2008). In fact, by discovering frequent patterns, the underlying temporal constraints, association rules, progress and changes over time, it is possible to characterize the behaviour of persons and objects and automate tasks such as activity monitoring, assistance and service adaptation (Rashidi and Cook, 2010).

Currently, there are many mathematical models for activity recognition, such as HMMs (Rabiner, 1989), Bayesian Networks (Oliver and Horvitz, 2005), Kalman Filters (Bodor et al., 2003) and Neural Networks (Bodor et al., 2003). Deep learning approaches on RGB video streams for activity recognition have also been introduced. This creates a system that improves and learns itself by updating the activity models incrementally over time (Hasan and Roy-Chowdhury, 2015).

Traditionally, most activity recognition work has focused on representing and learning the sequential and temporal characteristics in activity sequences. This has led to the widespread use of the HMM. In fact, in (Sung et al., 2011) HMM is employed with depth images to effectively recognize human activities. An HMM (Rabiner, 1989) is a finite set of states; each state is linked with a probability distribution. Transitions among these states are governed by a set of probabilities called transition probabilities. In a particular state a possible outcome or observation can be generated, according to the associated observation probability distribution. It is only the outcome, not the state that is visible to an external observer and therefore states are "hidden" to the outside, hence the name Hidden Markov Model. In earlier exploratory studies the HMM has shown good results thanks to their suitability to model sequential data, which is the case for monitoring human activities. Indeed, acceleration data are measured over time during physical human activities of a person and are therefore sequential over time. In (Coppola et al., 2016) an approach to activity recognition for indoor environments based on incremental modelling of long-term spatial and temporal context is presented. Even in (Coppola et al., 2015) the authors introduced a simple way to apply qualitative trajectory calculus to model 3D movements of the tracked human body using HMMs. HMMs combined with Gaussian Mixture Models (GMM) to model the combination of continuous joint positions over time for activity recognition was introduced in (Piyathilaka and Kodagoda, 2015).

In this paper, a method for ADLs recognition is proposed. In particular, we focus on using the HMM to facilitate the detection of anomalous sequences in a classical action sequence such as making a coffee.

3 DESIGN OF HMM STRUCTURE

Let $X = \{x_1, x_2, \ldots, x_n\}$ be a discrete finite activity space and $O = \{o_1, o_2, \ldots, o_m\}$ the observation space of a Hidden Markov Model (HMM) (Rabiner, 1989). Let T be the transition matrix of this HMM, with $T_{x,y}$ representing the probability of transitioning from activity $x \in X$ to activity $y \in X$, and $p_x(o)$ be the emission probability of observation $o \in O$ in activity $x \in X$.

We denote the probability that HMM trajectory follows the activity sequence s given the sequence of n observations, as:

$$P(X_{1:n} \in seq_n(s) | o_{1:n})$$

where $seq_n(s)$ is a set of all length n trajectories whose duration free sequence equals to s.

Finding the most probable activity sequence can be seen as a search problem that requires evaluation of probabilities of activity sequences. The Viterbi algorithm based on dynamic programming can be used to efficiently find the most probable trajectory. In fact, it makes use of the Markov property of an HMM (that the next state transition and symbol emission depend only upon the current state) to determine, in linear time with respect to the length of the emission sequence, the most likely path through the states of a model which might have generated a given sequence.

3.1 Head and hands detection algorithms

The main goal of this work is to classify different activities that people carry out during their daily life using an RGB-D camera in a top-view configuration. The idea is to extract from depth information the $3D$ position of the person for each frame. In particular, using the multi-level segmentation algorithm in (Liciotti et al., 2014), we can track the head and the hands of each person when these are visible. In fact, this algorithm intends to overcome the limitations of the single-level segmentation in the case of collisions among people in the same scene.

The multi-level segmentation algorithm is based on the idea that, normally, people head is on the higher level than the rest of the body. Thus, we can detect a person starting from its head. The minimum point of the blob that identifies a person, which is extracted from the depth image, corresponds to the head top point. When there is only one person in the scene,

the problem can be reduced to find the global minimum of the whole depth image. On the contrary, when more than one person is in the scene, we have to consider all the local minima. In fact, in case of collision, two people become a single blob and using binary segmentation it is not possible to recognise the two heads.

To solve this problem, we propose an algorithm that iterates the binary segmentation considering decreasing values of thresholds, until we obtain people profiles. Starting from the top-view, a set of contour levels is identified. For each level, the minimum point is stored and if a new local minimum higher than the previously stored is found, it will be added as the head point. When the loop ends, all the heads detected in the scene are recognised.

In a similar way, to find the hands $3D$ position, we apply again this algorithm to each person blob leaving out the upper part of person profile (head and shoulders) previously found.

3.2 ADLs Model

In this section, an ADLs model is described. It takes into account both the complexity of our data and the lack of a large amount of training data for learning purposes. If we abstract our problem of recognition of daily activities in the image to its simplest core, we can notice an equivalence between an activity and a hidden state of an HMM. This could be obtained with the design of a fully connected HMM and training the inherent state-transition probabilities from the labeled data. Regarding these ADLs as very heterogeneous and complex, the suggested equivalence between an activity and a hidden state cannot hold.

Information provided by head and hands detection algorithms can be used as input for a set of HMMs. Each of these recognise different actions sequence. After training the model, we consider an action sequence $s = \{s_1, s_2, \ldots, s_n\}$ and calculate its probability λ for the observation sequence $P(s|\lambda)$. Then we classify the action as the one which has the largest

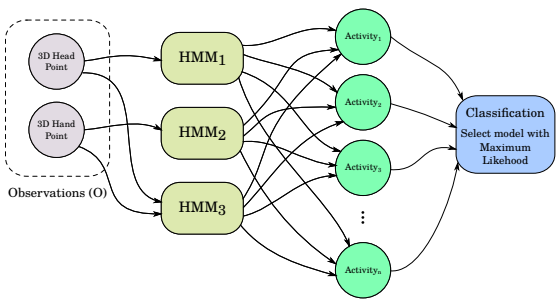


Figure 1: Block diagram of the recognition process.

Table 1: Number of observations for each HMMs (v: vertical layer, h: horizontal layer).

***3D* Points**	**# layers**	**# observations**
head	$v:8,\ h:8$	512
hands	$v:8,\ h:8$	512
head & hands	$v:8,\ h:8;\ v:8,\ h:8$	262144

posterior probability.

Figure 1 depicts the general scheme of the recognition process. In particular, we used three different HMMs, which have as observations 3D points of:

- the head (HMM_1);
- the hands (HMM_2);
- both head and hands together (HMM_3).

Table 1 indicates the number of vertical and horizontal layers used in the quantization step for each HMM and the total number of observations, after the resampling process.

Our set of actions includes:

- making a coffee;
- taking the kettle;
- making tea or taking sugar;
- opening the fridge;
- other activities performed in a kitchen environment.

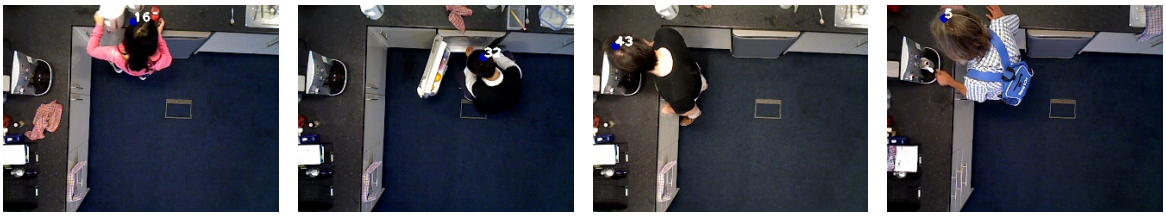
Finally, the classification module provides the action x_j that maximizes P_{HMM_i}. It is the HMM trajectory probability that follows the activity sequence s given the sequence of n observations, i.e.:

$$x_j = \arg\max_i P_{HMM_i}(X_{1:n} \in seq_n(s)|o_{1:n}) \qquad (1)$$

4 SETUP AND ACQUISITION

To evaluate the usefulness of our approach for activity recognition, we built a new dataset. The RADiAL dataset contains common daily activities such as making coffee, making tea, opening the fridge and using the kettle. The data were collected over a period of 5 days.

RADiAL also consists of random activities of each individual that can be performed in a kitchen environment, which are not similar to any other activity done before. The RGB-D camera was installed on the ceiling of L-CAS laboratory at approximately $4m$ above the floor. The camera was positioned above the surface which has to be analysed (Figure 3).



(a) (b) (c) (d)

Figure 2: Snapshots of RADiAL session registration.

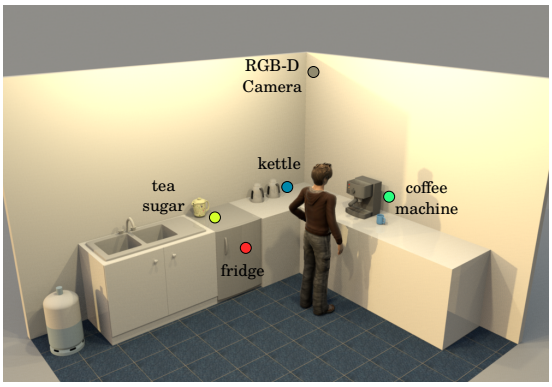


Figure 3: Reconstructed layout of the kitchenette where RGB-D camera is installed.

We chose an Asus Xtion Pro Live RGB-D camera for acquiring colour and depth information in an affordable and fast way. A mini PC for elaborating and storing the 3D points and colour frames was used.

4.1 RADiAL Dataset

The RADiAL dataset[1] was collected in an open-plan office of the Lincoln Centre for Autonomous Systems (L-CAS). The office consists of a kitchenette, resting area, lounge and 20 working places that are occupied by students and postdoctoral researchers. We installed a ceiling RGB-D camera (Figure 3) that took a snapshot (with dimensions of 320×240 pixels, Figure 2) of the kitchenette area every second for 5 days, and we hand-annotated activities of one of the researchers over time. Furthermore, the RADiAL dataset contains the 3D positions of the head and hands for each person with a minute-by-minute timeline of 5 different activities performed at the kitchen over the course of days. RADiAL contains 100 trials. Each trial includes the actions related to one person.

[1] http://vrai.dii.univpm.it/radial-dataset

5 EXPERIMENTAL RESULTS

In this section we present the experimental results obtained using our approach. An architecture to implement HMMs ADLs recognition is proposed. The architecture uses the 3D points extracted from the head and hands to classify different sequences of actions corresponding to some ADLs.

The standard algorithm for HMM training is the forward-backward, or Baum-Welch algorithm (Baum, 1972). Baum-Welch is an iterative algorithm that uses an iterative expectation/maximization process to find an HMM which is a local maximum in its likelihood to have generated a set of training observation sequences. This step is needed because the state paths are hidden, and the equations cannot be solved analytically.

In this study, the BaumWelch algorithm was employed to estimate a transition probability matrix and an observation emission matrix so that the model best fits the training dataset.

Since the discrete observation density is used in implementing HMMs, a Vector Quantization and clustering step is required to map the continuous observation in order to convert continuous data to discrete data.

A total of five models of activities were built using the method described in Subsection 3.2. The models were used to recognize activities in the RADiAL dataset. The five models correspond respectively to the activities “Other” (this action contains all the other activities perfomed in a kitchen environment), “Coffee” (making a coffee), “Kettle” (taking the kettle), “tea/sugar” (making tea or taking sugar), and “fridge” (opening the fridge). The results were obtained using two different validation techniques.

Below, the results are given at first for the head only (HMM_1), then the hands (HMM_2) and finally, the combination of both (HMM_3). The main goal is to gradually improve the activities recognition.

In the first case we applied a k-fold cross-validation approach (with $k = 5$) to test our HMM_1. The resulting confusion matrix is shown in Figure 4(a). As we can see in the confusion matrix

most of the actions are detected with high accuracy. Table 2 summarises the activity recognition results demonstrating the effectiveness and suitability of our approach.

Table 2: Classification Results Cross Validation HMM_1

	precision	**recall**	**f1-score**
other	0.73	0.57	0.64
coffee	0.67	0.80	0.73
kettle	0.60	0.70	0.65
tea/sugar	0.66	0.70	0.68
fridge	0.74	0.61	0.67
avg / total	0.68	0.68	0.68

The confusion matrix for HMM_2 is depicted in Figure 4(b). The activity recognition results, as reported in Table 3, prove the effectiveness and suitability in terms of precision, recall and f1-score.

Table 3: Classification Results Cross Validation HMM_2

	precision	**recall**	**f1-score**
other	0.89	0.70	0.79
coffee	0.69	0.83	0.75
kettle	0.47	0.58	0.52
tea/sugar	0.64	0.68	0.66
fridge	0.74	0.65	0.69
avg / total	0.73	0.71	0.71

The confusion matrix of HMM_3 for both the head and hands is shown in Figure 4(c). The results in Table 4 indicate an increase in the metrics for evaluating the performance of our approach.

Table 4: Classification Results Cross Validation HMM_3

	precision	**recall**	**f1-score**
other	0.93	0.76	0.84
coffee	0.76	0.87	0.81
kettle	0.58	0.70	0.63
tea/sugar	0.70	0.75	0.72
fridge	0.78	0.71	0.74
avg / total	0.78	0.77	0.77

In the second case, in order to demonstrate the suitability of this approach in a blind test, we split the dataset into two parts: 80 trials were used to train the model, the other 20 were used to test it. The aim of this experiment is to understand how our HMMs predict some activities. The main classification metrics precision, recall, f1-score and support are employed to evaluate the quality of predictions of our model. The results are reported in Tables 5 to 7.

Table 5 indicates the performance of the HMM_1 designed for the head. The confusion matrix for HMM_1 is also depicted in Figure 5(a). The activity recognition results of the overall activities demonstrate the effectiveness and suitability of our approach.

Table 5: Classification Results HMM_1

	precision	**recall**	**f1-score**	**support**
other	0.68	0.51	0.59	237
coffee	0.96	0.89	0.92	589
kettle	0.66	0.83	0.74	248
tea/sugar	0.75	0.87	0.81	636
fridge	0.90	0.64	0.75	233
avg / total	0.81	0.80	0.80	1943

Table 6 shows classification results of the HMM_2 designed for the hands. Classification metrics taken in exam overall decrease in performance, but there is an increase of the accuracy for the action "open the fridge". Probably, this is because the test set is not balanced. The confusion matrix for HMM_2 is also depicted in Figure 5(b).

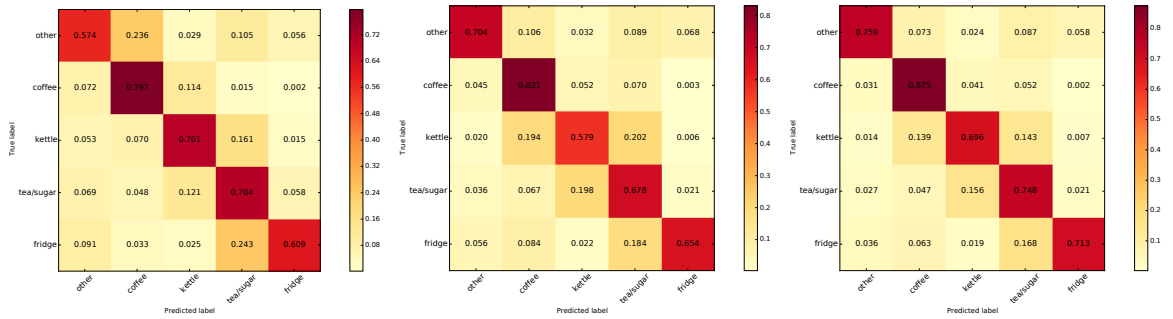
Table 6: Classification Results HMM_2

	precision	**recall**	**f1-score**	**support**
other	0.15	0.08	0.10	381
coffee	0.55	0.74	0.63	925
kettle	0.04	0.49	0.08	80
tea/sugar	0.63	0.27	0.38	1741
fridge	0.73	0.72	0.73	556
avg / total	0.56	0.44	0.46	3683

Table 7 shows classification results of the HMM_3 designed for both head and hands. Results in terms of precision are interesting. In fact, for some activities the precision is increased. The confusion matrix for HMM_3 is also shown in Figure 5(c).

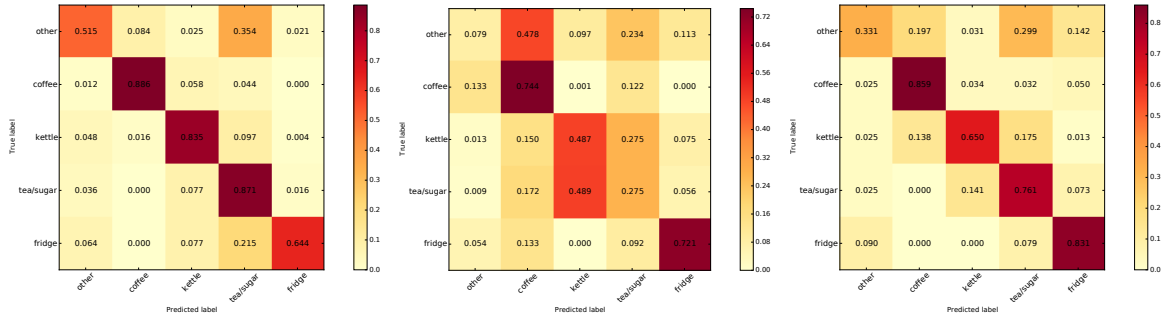
Table 7: Classification Results HMM_3

	precision	**recall**	**f1-score**	**support**
other	0.52	0.33	0.40	381
coffee	0.90	0.86	0.88	925
kettle	0.15	0.65	0.25	80
tea/sugar	0.87	0.76	0.81	1741
fridge	0.67	0.83	0.74	556
avg / total	0.79	0.75	0.76	3683



(a) HMM_1 (b) HMM_2 (c) HMM_3

Figure 4: k-fold cross-validation confusion matrices.



(a) HMM_1 (b) HMM_2 (c) HMM_3

Figure 5: Confusion matrices after splitted dataset in two parts: 80 trials are used to train the model, the other 20 are used to test it.

6 CONCLUSION

In this paper, we propose an automated RGB-D video analysis system that recognises human ADLs, related to classical daily actions that a person performs in the kitchen.

RADiAL, a dataset with RGB-D images and 3D position of each person for training as well as evaluating HMM, has been developed and made publicly available.

Action sequence recognition is then handled using a discriminative HMM. In particular, three different HMMs are used, which have as observations the 3D points of the head, hands and both together.

We conducted an evaluation of the performance of our approach on 5 activities performed in the RADiAL dataset. The experimental results demonstrate the effectiveness and suitability of our model that achieves high accuracy and performs well, without having to rely on the data annotation required in the other existing approaches.

Further investigation will be devoted to extend our approach to select human joints that provide the most informative spatio-temporal relations for ADLs classification.

Future work includes performing the acquisition by other more accurate RGB-D cameras, such as a TOF camera.

Moreover, it would be interesting to evaluate both color and depth images in a way that does not decrease the performance of the system when the color image is being affected by changes in pose and/or illumination.

In the field of assistive technology, the long term goal of this work is to develop a mobile robot that searches for the best location to observe and successfully recognise ADLs in domestic environments.

ACKNOWLEDGEMENTS

The authors would like to thank L-CAS Team for the support during the dataset video acquisition.

REFERENCES

Ali, R., ElHelw, M., Atallah, L., Lo, B., and Yang, G.-Z. (2008). Pattern mining for routine behaviour discovery in pervasive healthcare environments. In *2008 International Conference on Information Technology and Applications in Biomedicine*, pages 241–244. IEEE.

Baum, L. E. (1972). An equality and associated maximization technique in statistical estimation for probabilistic functions of markov processes. *Inequalities*, 3:1–8.

Bodor, R., Jackson, B., and Papanikolopoulos, N. (2003). Vision-based human tracking and activity recognition. In *Proc. of the 11th Mediterranean Conf. on Control and Automation*, volume 1. Citeseer.

Coppola, C., Krajnık, T., Duckett, T., and Bellotto, N. (2016). Learning temporal context for activity recognition. In *ECAI 2016: 22nd European Conference on Artificial Intelligence, 29 August-2 September 2016, The Hague, The Netherlands-Including Prestigious Applications of Artificial Intelligence (PAIS 2016)*, volume 285, page 107. IOS Press.

Coppola, C., Martinez Mozos, O., Bellotto, N., et al. (2015). Applying a 3d qualitative trajectory calculus to human action recognition using depth cameras. In *IEEE/RSJ IROS Workshop on Assistance and Service Robotics in a Human Environment*.

Dartigues, J. (2005). [methodological problems in clinical and epidemiological research on ageing]. *Revue d'épidémiologie et de santé publique*, 53(3):243–249.

Gupta, A., Srinivasan, P., Shi, J., and Davis, L. S. (2009). Understanding videos, constructing plots learning a visually grounded storyline model from annotated videos. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 2012–2019. IEEE.

Hasan, M. and Roy-Chowdhury, A. K. (2015). A continuous learning framework for activity recognition using deep hybrid feature models. *IEEE Transactions on Multimedia*, 17(11):1909–1922.

Hoai, M. and De la Torre, F. (2014). Max-margin early event detectors. *International Journal of Computer Vision*, 107(2):191–202.

Jiang, Y. and Saxena, A. (2013). Infinite latent conditional random fields for modeling environments through humans. In *Robotics: Science and Systems*.

Koppula, H. S., Gupta, R., and Saxena, A. (2013). Learning human activities and object affordances from rgb-d videos. *The International Journal of Robotics Research*, 32(8):951–970.

Liciotti, D., Contigiani, M., Frontoni, E., Mancini, A., Zingaretti, P., and Placidi, V. (2014). Shopper analytics: a customer activity recognition system using a distributed rgb-d camera network. In *International Workshop on Video Analytics for Audience Measurement in Retail and Digital Signage*, pages 146–157. Springer International Publishing.

Liciotti, D., Massi, G., Frontoni, E., Mancini, A., and Zingaretti, P. (2015). Human activity analysis for in-home fall risk assessment. In *2015 IEEE International Conference on Communication Workshop (ICCW)*, pages 284–289. IEEE.

Liu, J., Ali, S., and Shah, M. (2008). Recognizing human actions using multiple features. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE.

Maji, S., Bourdev, L., and Malik, J. (2011). Action recognition from a distributed representation of pose and appearance. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 3177–3184. IEEE.

Nait-Charif, H. and McKenna, S. J. (2004). Activity summarisation and fall detection in a supportive home environment. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 4, pages 323–326. IEEE.

Ning, H., Han, T. X., Walther, D. B., Liu, M., and Huang, T. S. (2009). Hierarchical space-time model enabling efficient search for human actions. *IEEE Transactions on Circuits and Systems for Video Technology*, 19(6):808–820.

Oliver, N. and Horvitz, E. (2005). A comparison of hmms and dynamic bayesian networks for recognizing office activities. In *International conference on user modeling*, pages 199–209. Springer.

Pirsiavash, H. and Ramanan, D. (2012). Detecting activities of daily living in first-person camera views. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2847–2854. IEEE.

Piyathilaka, L. and Kodagoda, S. (2015). Human activity recognition for domestic robots. In *Field and Service Robotics*, pages 395–408. Springer.

Prest, A., Schmid, C., and Ferrari, V. (2012). Weakly supervised learning of interactions between humans and objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(3):601–614.

Rabiner, L. R. (1989). A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286.

Rashidi, P. and Cook, D. J. (2010). An adaptive sensor mining framework for pervasive computing applications. In *Knowledge Discovery from Sensor Data*, pages 154–174. Springer.

Ryoo, M. S. (2011). Human activity prediction: Early recognition of ongoing activities from streaming videos. In *2011 International Conference on Computer Vision*, pages 1036–1043. IEEE.

Sturari, M., Liciotti, D., Pierdicca, R., Frontoni, E., Mancini, A., Contigiani, M., and Zingaretti, P. (2016). Robust and affordable retail customer profiling by vision and radio beacon sensor fusion. *Pattern Recognition Letters*.

Sung, J., Ponce, C., Selman, B., and Saxena, A. (2011). Human activity detection from rgbd images. *plan, activity, and intent recognition*, 64.

Wu, J., Osuntogun, A., Choudhury, T., Philipose, M., and Rehg, J. M. (2007). A scalable approach to activity recognition based on object use. In *2007 IEEE 11th International Conference on Computer Vision*, pages 1–8. IEEE.